

# CLASSIFICATION BASED ON GEOGRAPHICAL INFORMATION ONTOLOGY

Hong Wang, Lin Li

School of Resource and Environmental Science, Wuhan University, Wuhan City, China-j-wanghong@163.com

**KEY WORDS:** Geographical information, Ontology, Property, Classification

## ABSTRACT:

Ontology, which derives from philosophy, is a discipline of Philosophy that deals with existence. In recent years, it has been widely used in artificial intelligence and computer science, mainly representing, sharing and reusing knowledge. So it's significant to introduce ontology into Geographic Information Science for the implementation of the geographical information representation, sharing and reuse. Based on analyses of geographical information taxonomy, after explaining the concept of ontology, this paper expatiates a method of building Geographical Information Ontology and a process of classifying geographical information concepts. In classification, it places emphasis on using ontology for method, starting from property of concepts, analysing the property of the concepts sufficiently, working out the priorities of ontological properties, and carries out the classification of concepts based on ontological properties.

## 1. INTRODUCTION

Ontology is a discipline of Philosophy that deals with what it is, with the kinds and structures of objects, properties, and other aspects of reality. In recent years, Research on ontology is becoming widespread in artificial intelligence and computer science. These successful applications make solid foundations for applying ontology to geographical information science. The history of early ontology applied to geographical information can trace back to the expert symposium on the 21<sup>st</sup> research motive of NCGIA in USA, 1996, in which Mark brought forward that ontology should be included as a research topic of NCGIA, and made an incomplete draft of geographical ontology (Mark, 1997); later Smith and Mark (1998, 1999) systematically studied categories of geographical information on the aspect of ontology, and suggested that geographical categories should be studied by means of mereology, topology and qualitative geometry. Compared with this studies, Our domestic study starts less later, such as follows: Jianbang He (2003) put forward an ontological model of the classification and coding of geographical information based on both macro and micro levels in the classification and coding of resource and environment information; Wei Song (2004) in his book "A First Step Towards the Semantic Web" made a detailed account of ontology and ontological description languages, In the end, he gave the concept, methodology of building Geographical Information Ontology as an application examples. More and more researchers (Haide Zhang, 2002) have been working in this domain.

It is not difficult to see that the special focus on Geographical Information Ontology reflects not only urgent demands of research on this aspect, but also a law of things development, that is to say, we concern more about the nature of things, i.e. "what is it" or the ontology of things, rather than about the surface of things in the development of things. Tracing back to the nature of things, we can find out profound and inbeing of this world.

## 2. THE MEANING OF ONTOLOGY

Various research communities gave different explanations and definitions of ontology with their purposes, the most representative is the definition made by Gruber (1993) and Guarino (1998). Although different researchers hold different descriptions of ontology, they all consider ontology as a semantic basis of communication (dialog, interoperation, sharing, etc.) between different main bodies (human, agent, machine, etc.) in a certain domain (a specific domain or a wider domain). It means that ontology can provide a specifically defined glossary as the commonsense among users to describe concepts and the relations between them.

Ontology includes concepts, taxonomy and algebra, and so on. In these theories, the most representative is the in-depth study on taxonomy made by Guarino and his colleagues (Guarino, 2000; Welty, 2001). They proposed a series of practicable theories to guide the taxonomy based on concepts, properties of concepts, relations between concepts, and taxonomy. It is obvious that taxonomy is an important part of ontology.

Ontological taxonomy tries to specifically represent complex semantics of concepts, concepts, properties of concepts, variables, and relations between constraints, mainly including the following (Wiley, 2003):

- Classes (general things) in many domains of interest
- Instances (particular things)
- Relationships among those things
- Properties (and property values) of those things
- Functions and processes involving those things
- Constraints on and rules involving those things

Ontological taxonomy includes above relations. In the classification, from the point of concepts, ontology uses the basic taxonomic subclasses of hierarchies as the skeleton of ontology. But muscles and organs are added in ontology—in the form of additional relations, properties/attributes, property values. So taxonomies provide basic structure for the information space, and ontology flesh it out.

### 3. CURRENT CLASSIFICATION OF GEOGRAPHICAL INFORMATION

The current classification of geographical information adopts the linear classifying method (Shuoben Bi, 2003) for feature classification. This method classifies initial objects for classification to several hierarchical catalogues based on several selected properties or characteristics, and arranges them to a hierarchical classification system. Its advantage is as follows: classification based on experience, large volume, and good hierarchy, and so on. However, for extra subject factors are added to the classification, there are inevitably following problems:

- (1) Classification indexes of features may differ from each other. Sometimes for convenience, features at the same level are not classified according to the same standard, leading to disaccords in descriptions of features (GB/T 13923-1992).
- (2) Concepts of features are not fully reflected in this classification but often expressed later in symbol specification.
- (3) Only the common up-down relation between features can be expressed. This classification can't be applied when a certain feature may belong to other different features based on concepts and properties.
- (4) Properties of features can't be specifically expressed in this classification.

Based on above analyses, this paper refers to vocabulary and notable foreign and domestic ontology (Hownet, WordNet), and makes a re-classification of geographical features from the point of ontology view, which bases itself on concepts and conceptual properties of geographical information, and takes characteristics of our national language and geographical features.

## 4. BUILDING GEOGRAPHICAL INFORMATION ONTOLOGY

### 4.1 Significance of Geographical Information Ontology

From the representation form, Geographical Information Ontology is similar to the classification code of geographical information and standard geographical information terms list. However, in Geographical Information Ontology, relations between concepts are completely and fully reflected by using corresponding technologies such as predications, logic and so on. It is the reflection of these relations that classification systems based on ontology can agree with each other on semantic aspects. Ontology has made efforts to make itself computer-understandable since its birth, so it needs more special technologies to support it from the aspect of representation form. For example, exact formal languages, syntax, and specified semantics are used in the representation of ontology. Ontology can provide theoretical and technological support for the standardization of classification encoding of geographical information.

Using for reference of six elements(Naing,2002) of ontology, We give the formal definition of Geographic Information Ontology as follows:  $GIO = \{ C, A^c, R, A^R, H, X \}$ , where  $C$  represents a set of concepts;  $A^c$  represents a collection of attribute sets, one for each concept;  $R$  represents a set of relationships;  $A^R$  represents a collection of attribute sets, one

for each relationship;  $H$  represents a concept hierarchy; and  $X$  represents a set of axioms.

In GIO, its vocabulary comes from Chinese, so some classification of natural features can refer to Chinese Thesaurus «Cilin» and Hownet. However, it differs from the former two. GIO takes in expert knowledge in geographical information science as its content, which emphases that concepts in domains are basic units of semantics and each concept should be represented through relations between concepts and properties; «Cilin» is just created from the aspect of Chinese, which reflects very simple synonymical and upper-down relations between vocabularies in our daily languages. Every word sense is represented by the combination of several sememes and a sememe is a basic semantic unit that is indivisible in Hownet. GIO is human's understanding of things and laws in nature, and it reflects relations among things, which itself can be independent of any language.

GIO obtains concepts and relations among concepts, which stored in the form of sets of GIO concepts (sets of all concepts), sets of GIO properties (properties corresponding to concepts) and sets of GIO relations (relations between concepts).

### 4.2 Methodologies for Building Geographical Information Ontology

Current methodologies for the development of knowledge ontology can be grouped into two categories. One is the top-down methodology, by which we should first construct macro and abstract framework of conceptual systems from a top level, and then according to application demands, specify and improve knowledge ontology step by step. The other is the bottom-up methodology, by which we should first construct conceptual models and formal representation based on universal ontological languages from specific applications, and later extend ontology through practicable applications.

The bottom-up methodology has its own advantages, one of which is that we can describe our defined ontology in detail according to specific applications. But now that it aims at specific applications, it unavoidably depends on specific domains, concepts of tasks and relations between concepts. Meanwhile, Geographical Information Ontology is domain-oriented knowledge ontology, which makes it necessary to construct abstract conceptual systems first in the development process of Geographical Information Ontology. Thus, we adopt the top-down methodology to construct Geographical Information Ontology (GIO for short), refer to corresponding documents and study on the classification of geographical information by both domestic and overseas domain experts, also introduce the methodology to construct ontology used in knowledge engineering (Natalya F, 2001), and determine the method of constructing GIO as flows:

1. Determine the domain and scope of the ontology
2. Enumerate concepts and properties
3. Define relations between concepts by properties
4. Determine property variables
5. Create instances
6. Describe GIO in formal language
7. Create embryo of GIO ontology
8. Evaluate ontology
9. Create formal ontology

It is easy to infer that as we specifying concepts and their properties, the classification of concepts is the first problem in the process of constructing GIO.

## 5. CLASSIFICATION OF GEOGRAPHICAL INFORMATION BASED ON ONTOLOGY

### 5.1 Properties and Classification

In the classification of information systems, properties are mainly used to represent knowledge classification. Different properties play different roles in the representation of knowledge classification. Some properties are absolutely unnecessary, and knowledge discovery can't be affected if these properties are missing; some properties are absolutely necessary, so if they are ignored, knowledge discovery is surely affected; some are relatively necessary, and it affects knowledge discovery only when it matches all absolutely necessary properties. Therefore, properties take an important role in the knowledge classification.

Aristotle's theories on ontology consider that (Zisong Wang, 1997) properties can't independently exist if separated from ontology. He specifies that any single (particular) ontology is destined to occur and die, even when it exists, its all properties will change ceaselessly, so what we call "constant in changes" is just a relative concept. It is to be specified that ontology (different from epistemology) be about thoughts of existence, which thinks that an object doesn't change itself, but only its properties. It is regarded that properties of matter reflect and follow natural laws, which is necessary and universal by nature.

Jianbang He (2003) thinks that property is the abstraction of qualitative and quantitative characteristics of entities, which defines essential properties and semantic relations (non-spatial relations) of resource and environment entities. He thinks that property is the basis of classification encoding of resource and environment information. As for entities, they are sets of the same property description objects (things or phenomena).

Properties of things mainly include intrinsic properties and extrinsic properties. Intrinsic properties, possessed by things themselves, are natural properties of things, and they can't be changed by human mind; extrinsic properties (such as names, possession, etc.) have social properties, so their representation forms may be different when applied to different objects or different periods. Obviously intrinsic properties should have higher priority than extrinsic properties. This paper herein sets focus on properties of geographical information concepts, searches for ontological properties of concepts in change, and specifies their priority, thus implements the classification of geographical information concepts.

### 5.2 Ontological Properties Of Physical Geographical Features

There are following issues we should take into account before the classification of geographical information concepts: (1) As for the same objective geographical entity, whether all departments adopt the same name, properties, and property name description, namely whether geographical entities have standard conceptual model; (2) As the same geographical entity in different departments have the same applications, whether or not these applications have the same definitions. Aiming at these two cases, we define concepts like this: if

different departments name about concept and property differently, we choose a common name, while other names are in the form of aliases; this classification is based on basic geographical information, so it can't include all departments, and we only choose some often-used properties of basic geographical information, and other departments can be accessed by transferring with the help of codes. Geographical features mainly include physical geographical features and social features. As a great deal of subjective factors exist in the classification of social features, this paper takes physical geographical features for example, and classifies them based on characteristics of properties from the point of ontology view.

Borgo (et al, 1996) and Guarino (1997) put forward "ontological strata" in their study of building top-level ontology. They think that an object has an identity criterion, objects can be categorized by different identity criterion, and this results in strata called ontological strata. Frank (2001) put forward a theory of ontological tier suited GIS and other geographical entities, in which different classification rules suit different layers. In the light of above experience, we obtain priority ordering of properties (Table 1) in the process of classification of physical geographical features.

Macroscopic view	Mereological (an amount of matter) tier
	Morphological tier (a cubic block)
Microcosmic view	Real-time tier
	Morphological tier
	Functional tier
	Social tier

Table 1. Property Priority of physical geographical features

On the macro aspect, we classify all things existing in nature from both mereological and morphological view.

#### (1) Mereological tier

From the point of semantics of spatial information, theory of mereology should be included first in the classification of ontological properties based on concepts, in which this theory is used to describe relations between parts and wholes. As for this mereological tier, it is mainly used to represent "what is it". So according to this, physical geographical features can be divided into features such as water, vegetation, soil and so on.

#### (2) Morphological tier

Answering the question of "what is it", we want to know what kind of spatial configuration it forms. Is it fluid or static, regular or irregular? Spatial configuration properties of concepts become a second important property in the ontological classification.

We can use above methods to classify our familiar physical phenomena. Moreover, we can also obtain priority list of these concepts with the help of «Cilin» and Hownet to further test if our classification corresponds to our general thinking process. However, this kind of classification of geographical features can't fully meet our demands, still we need to know exactly the composition of concepts. Therefore, we must classify from the micro point of view.

On the micro aspect, classification is made according to intrinsic properties of things, involving following primary properties:

#### (1) Real-time tier

It is known that properties of physical geographical features change continuously with time. For example, river can be cut by outside environments, and its shape would be changed from a double-line river to a small lake, and finally to a billabong. Therefore, it is necessary to keep this property possessed by a sole thing rather than another thing. And so temporal nature expressed during the change of features is the first property that should be considered in our property classification. This property is the most correlatives ontological property of a feature.

(2) Morphological tier

The size of a feature is another important property in the classification of physical geographical features, for different sizes determining different types of physical geographical features. In geographical space, classification often relates to dimension or scale.

Following that, we can further classify concrete geographical features according to their functional and social properties.

There surely are some disadvantages if we just apply above rules and standard classification of objects in our daily world, so we need to further summarize additional rules with the integration of characteristics of geographical objects and geographical phenomena. Moreover, in the process of classification, there is no hard rule for the number of direct subclasses that a class should have. However, many well-structured ontology have direct subclasses about two to a dozen. Therefore, there are following guiding principles for the GIO classification (Natalya F. , 2001).

If a class has only one direct subclass, there may be a modelling problem or the ontology is not complete, or if there are more than a dozen subclasses for a given class, then additional intermediate categories may be necessary.

**6. CLASSIFICATION OF PHYSICAL GEOGRAPHICAL FEATURES BASED ON ONTOLOGICAL PROPERTIES**

Physical geographical features are composed of hydrologic features, physiognomy, soil, vegetation and so on. Among them, Hydrologic features can be considered as one of the most typical physical geographical features. This paper makes a classification of this feature on the basis of its ontological properties.

According to morphological tier of hydrologic features, it can be classified as rivers, lakes, ponds, oceans, etc. In Hownet, the degree of similarity between them and hydrologic features is all 0.6; in «Cilin», they are at the same layer of the dendriform structure. We can see that the classification based on morphological tier agrees to human’s way of understanding things, it approves that the result is correct according to this method. Now we take lakes for example to illustrate the classification process in detail.

(1) Lake first should be water body, so it’s necessary to classify lake by the period water in lakes. In this way, it can be divided as perennial water, seasonal lake, and dry lake that only have water because of rain falling or snow melting. Water’s real-time property can be considered as the first ontological properties in the classification of lakes. Moreover,

other properties of lakes are as follows: the degree of lake mineralization, area, ascription and name.

(2) After the determination of types of lake, the size of Morphology of lake, area, is the second important property of lake. Lake may become to ponds if their areas get smaller than a certain constant. Because of this important property, geographical entities differ from the rest entities.

(3) After types of an entity are decided, from the ontological point of view, we analyse properties left and can find that the quality of features is an important issue. As for lakes, its quality mineralization is the third important property.

(4) Lastly, let us focus on two extrinsic properties, name and ascription. Compared with name, ascription tends to change more quickly during the same period.

Finally, we can list priority level of ontological properties of lake as follows in Table 2:

Attribute	Priority
Water Duration	1
Area	2
Degree of mineralization	3
Name	4
Ascription	5

Table 2. Priority level of ontological properties of lake

By this methodology, we can implement classification of hydrologic features, thus realizing the conceptualisation of features. However, because Chinese is a complex language and people have different habits while using it, the result of classification still needs to be rebuilt. Taking lake for example, in fact, the area property gains less priority than degree of mineralization. This can be explained in two aspects: on one hand, current topographic maps are all generalized, and so in this way they have classified specifically features by size; on the other hand, as computers already can handle area property quantitatively rather than qualitatively, we tend to accept quantitatively-handled results. Considering that physical geographical feature other than lakes may possess different representation from lakes, in our classification it is quite necessary to further process properties according to characteristics of concrete features.

**7. CONCLUSIONS**

This paper expatiates the process of classification based on ontological properties with concepts and illustrates it with physical geographic feature. Classification based on ontology can effectively express the meaning of our defined terms, and it is explicit and objective. Concepts described in the form of natural languages or instances, can also be consistent. Meanwhile, people can extend the existing vocabulary to new terms, without changing original definitions. Ontology deals with knowledge and hierarchy, has nothing to do with specific symbol encoding, this decreases the difficulty of code post-processing. However, building ontology is a complex process, which involve knowledge not only in professional domain, but also in fields such as linguistics, computer science and so on. Taking into account the complexity of geographical features and richness of Chinese language, how to search for

ontological properties of things in daedal world and lead our study from qualitative descriptions to quantitative descriptions, all these require our further deep research.

## REFERENCES

- Andrew U.Frank, 2001. Tiers of ontology and consistency constrains in geographical information systems. *Geographical Information Science*, Vol (15), pp.667-678.
- Borgo, S., Guarino, N., and Masolo, C., 1996. Stratified Ontology: the Case of Physical Objects. In: *Proceedings of ECAI-96 Workshop on Ontological Engineering*. Budapest.
- Dehai ZHANG, 2002. *Geographic Knowledge Acquisition and Analysis of Country and Region in NKI*, the master's thesis's of Yunna Normal University.
- GB/T 13923-92, 1992 Classify and codes for the national land information.
- Guarino, N., 1997.some Organizing Principles For A Unified Top-level Ontology. In: *Working Notes of AAAI Spring Symposium on Ontological Engineering*, Stanford, pp.57-63.
- Guarino, N., 1998. *Formal Ontology and Information Systems*. Proceeding of FOIS'98, Trenton, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp.3-15.
- Guarino, N., C Welty, 2000. A formal ontology of properties, in: R. Dieng (Ed.), *Knowledge Engineering and Knowledge Management*, Springer Verlag LNCS, Berlin, Germany.
- HowNet Knowledge, <http://www.keenage.com/>
- Jianbang HE, Xintong LI, et al, 2003. Research on Classification and Coding of Resources and Environment Information and Its Association with Ontology, *Geomatics World*, vol (1), no.5, pp.6-11.
- Jianbang HE, Xintong LI, 2002. Research on the Ontology of Geographic Information Classification and Coding, *Geography and Territorial Research*, vol, 18(3), pp.6-11.
- Mark, D.M et al., 1997. *Formal Models of Commonsense Geographic Worlds*. Report on the specialist Meeting of Research Initiative 21. NCGIA Technical Report 97-2.
- Mark, D. Max Egenhofer, et al., 2000. *Ontological Foundations for Geographic Information Science*. University Consortium for Geographic Information Science. *Emerging Themes in GIScience Research 2000*, White Paper. (<http://www.ucgis.org/emerging/>)
- Michael C. Daconta, Leo J. Obrst, Kevin T. Smith, 2003. *The Semantic Web: A Guide to the Future of XML Web Services and Knowledge Management*. John Wiley & Sons Inc.
- Naing, M.M., Lim, E. P., Goh, D.H, 2002. Ontology-based web annotation framework for hyperlink structures. *Proceedings of the International Workshop on Data Semantics in Web Information Systems*, Singapore, pp. 184-193.
- Natalya F. Noy, 2001 Deborah L. McGuinness [http://protege.stanford.edu/publications/ontology\\_development/ontology101-noy-mcguinness.html](http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html).
- Shuoben BI, Qiao WANG, Xiuhua XU, 2003. *The Principle and Method of Geographic Information System software engineering*. Beijing: Science Press.
- Smith, B. & Mark, D. M. ,1998. Ontology and geographic kinds. *Proceedings of SDH ' 98*, Vancouver, Canada.
- Smith, B. &Mark, D. M.,1999. *Ontology with Human Subject Testing*, *American Journal of Economics and Sociology* 58/2.
- Welty C, Guarino N., 2001. Supporting Ontological Analysis of Taxonomic Relationships. *Data and Knowledge Engineering*, 39(1), pp. 51-74.
- Wei SONG, Ming ZHANG, 2004. *A First Step Towards the Semantic Web*, Higher Education Press.
- Zisong WANG, 1997.*The Aristotle's Ontology*. People's Publishing House.